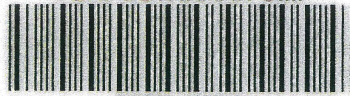


PERPUSTAKAAN UMP



0000117618


RESOLUSI ANAFORA ARTIKEL BAHASA MELAYU BERASASKAN  
PENGETAHUAN TERHAD DAN KELAS SEMANTIK

NOORHUZAIMI@KARIMAH MOHD NOOR

TESIS YANG DIKEMUKAKAN UNTUK MEMPEROLEH IJAZAH  
DOKTOR FALSAFAH

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT  
UNIVERSITI KEBANGSAAN MALAYSIA  
BANGI

2016

PERPUSTAKAAN  UNIVERSITI MALAYSIA PAHANG	
No. Perolehan <b>117618</b>	No. Panggilan PL 5104 -N66 2016 y 1111
Tarikh <b>03 APR 2017</b>	

**PENGAKUAN**

Saya akui karya ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang tiap-tiap satunya telah saya jelaskan sumbernya.

19 OGOS 2016

NOORHUZAIMI@KARIMAH MOHD NOOR  
P47842

## PENGHARGAAN

Segala puji ke hadrat Allah S.W.T yang maha menguasai sekalian alam, selawat dan salam buat junjungan mulia, Rasulullah S.A.W., ahli keluarga baginda, para sahabat serta para tabiin. Alhamdulillah hirrabbil'amin, dengan kerendahan hati saya merafakkan kesyukuran ke hadrat Ilahi di atas limpahan kasih sayang dan rahmat-Nya saya berpeluang menyiapkan tesis penyelidikan Ph.D ini. Jutaan terima kasih diucapkan seikhlasnya kepada penyelia utama Prof. Dr. Shahrul Azman Noah yang telah banyak memberikan sokongan, bimbingan, nasihat yang berguna dan bersabar sepanjang perjalanan saya menyempurnakan kajian ini. Tidak lupa ucapan terima kasih buat kedua-dua penyelia bersama saya iaitu Prof. Madya Dr. Mohd Juzaidin Ab Aziz dan Prof. Madya Dr. Mohd Pouzi Hamzah ke atas pandangan dan dorongan kepada saya. Segala jasa dan budi yang ditaburkan pasti akan saya kenang dan ingati seumur hidup saya. Hanya iringan doa yang mampu dipanjatkan, semoga dibalas dengan seribu kebaikan oleh Allah S.W.T.

Penghargaan juga ditujukan kepada Universiti Malaysia Pahang (UMP) dan Kementerian Pengajian Tinggi (KPT) yang telah memberikan peluang kepada saya untuk menyambung pengajian secara sepenuh masa dan menaja pengajian saya. Tidak dilupakan juga kepada pihak-pihak yang terlibat bersama-sama membantu dan memberikan idea serta sokongan untuk menjayakan kajian ini sama ada secara langsung atau tidak langsung.

Akhir kata, segulung penghargaan kasih sayang dijulung buat suami tercinta, Kasmadi Mahd Rashid, yang disanjung dan dihormati kedua-dua ibu bapa, Hajah Nik Hasnah Haji Nik Mahmood dan Haji Mohd Noor Mamat serta ibu mertua Siti Mas dan Allahyarham ayah mertua Rashid yang tidak pernah jemu berdoa, menasihati, bersabar dan berkorban untuk memberikan ruang kepada saya untuk fokus dalam kajian. Terima kasih juga ditujukan buat adik-adik tersayang, ahli keluarga, teman-teman seperjuangan, kakitangan Fakulti Sistem Komputer & Kejuruteraan Perisian (FSKKP) di atas segala bantuan dan sokongan yang diberikan terutama group KRONI.

## ABSTRAK

Resolusi anafora (RA) merupakan suatu proses penyelesaian rujukan yang melibatkan kata ganti nama (KGN). Ia merupakan fenomena yang berlaku dalam setiap bahasa dan memerlukan kepakaran manusia atau petua tertentu dalam menyelesaikannya. RA merupakan antara proses yang mampu meningkatkan kemampuan aplikasi pemprosesan bahasa tabii seperti sistem soal-jawab, perlombongan teks, peringkasan dokumen, dan pengekstrakan maklumat. Terdapat banyak kajian dilaksanakan oleh pengkaji terdahulu namun ianya hanya diselesaikan untuk bahasa tertentu seperti bahasa Inggeris, Jepun, dan Norwegian. Berdasarkan kajian yang dibuat hampir tiada atau sedikit sahaja kajian RA yang dibuat dalam bahasa Melayu (BM). Matlamat penyelidikan ini adalah untuk menyelesaikan fenomena rujukan KGN bagi bahasa Melayu menggunakan pengetahuan terhad dan model pelabelan kelas semantik. Bagi mencapai matlamat ini, satu rangka kerja resolusi untuk BM telah dihasilkan sebagai panduan dalam menyelesaikan masalah rujukan ini. Sementara itu, penentuan jenis KGN *nya* juga diselesaikan dengan mengguna petua dan jujukan perkataan setara dan perkataan saringan hasil daripada model pelabelan kelas semantik. Proses ini perlu kerana penggunaan KGN *nya* merupakan KGN yang paling tinggi digunakan dalam artikel BM iaitu sebanyak 68% berbanding dengan KGN lain yang bergantung kepada sosiologi entiti yang dirujuk. Penentuan calon entiti yang dirujuk atau anteseden juga merupakan proses yang penting dan perlu dititik beratkan. Antara calon anteseden tersebut yang perlu diambil kira adalah kata nama khas (KNK) dan kata nama am (KNA). Dalam menentu KNK sebagai calon anteseden dua proses perlu dilakukan iaitu memproses pengecaman KNK yang mengandungi perkataan '*dan*' dan simbol iaitu *koma* (,) dan proses yang kedua penentuan kelas semantik ke atas calon tersebut bagi menentukan sosiologi entiti nama tersebut. Penggunaan gazetir sebahagian KNK digunakan. Penilaian ke atas pelbagai bentuk KNK dijalankan dengan menggunakan 60 artikel BM. Perbandingan keputusan yang dihasilkan dengan data yang telah ditanda oleh pakar BM dilakukan bagi mendapat hasil ketepatan dan dapatan semula. Keputusan menunjukkan bahawa pengecaman dapat dilakukan dengan nilai purata ketepatan berjumlah 85% dan purata dapatan semula bernilai 90%. RA dengan mengguna pengetahuan terhad yang diadaptasi dalam kerangka kerja yang dibina memberi keputusan 18.79% kadar kejayaan lebih baik dibandingkan daripada hasil yang didapati jika menggunakan pendekatan generik Mitkov dan Lappin.

## ANAPHORA RESOLUTION FOR MALAY ARTICLE USING KNOWLEDGE POOR AND SEMANTIC CLASS

### ABSTRACT

Anaphora resolution (AR) is a process to resolve reference entity of pronoun anaphora. It is a phenomenon that occur in every languages and requires human experts or specific rules in order to resolve it. AR able to improve language processing applications such as question-answering, text mining, document summarizations, and information extraction. There has been various research carried out on AR, but the majority of them were meant for languages such as English, Japanese and Norwegian. Very few and almost no research effort have been focussed on AR for Malay language. Therefore, the aim of this research is to resolve the phenomena of AR for Malay text by using knowledge poor approach and semantic class labelling model. In order to achieve the aim, a framework of the Malay AR has been developed as a guide to solve this phenomenon in Malay language. Meanwhile, the process to determine the type of usage for pronoun *nya* has been solved by using a set of rules, a set of similar words, and word filtering that has been generate from semantic class labelling model. This process is important because the use of pronoun *nya* in Malay text is the highest, amounting to 68% as compared to other pronouns that mostly depend on the sociological status of referring entity or antecedent. The antecedent candidate determination is an important process that should be considered. The antecedent candidates can be in the form of proper noun or nouns. In order to determine proper nouns as suitable candidates, two main processes need to be done: (1) the entity recognition for proper noun that has the word '*dan*' and *comma* symbol (,); and (2) the process to determine the semantic label for each retrieved candidate in order to determine their sociological status. The research used part of the name gazetteers for people, organization, location and position. Testing has been conducted on 60 Malay articles with different classes of proper nouns. The results were compared with the benchmark data tagged by a Malay linguist. The result shows an average precision and recall values of 85% and 90% respectively. The proposed framework of AR by using knowledge poor approach for Malay text shows increased success rate by 18.79% as compared to the generic approach proposed by Mitkov and Lappin.

## KANDUNGAN

		<b>Halaman</b>
<b>PENGAKUAN</b>		<b>ii</b>
<b>PENGHARGAAN</b>		<b>iii</b>
<b>ABSTRAK</b>		<b>iv</b>
<b>ABSTRACT</b>		<b>v</b>
<b>KANDUNGAN</b>		<b>vi</b>
<b>SENARAI JADUAL</b>		<b>x</b>
<b>SENARAI RAJAH</b>		<b>xi</b>
<b>SENARAI SINGKATAN</b>		<b>xiii</b>
<b>SENARAI ISTILAH</b>		<b>xiv</b>
<b>BAB I</b>	<b>Pengenalan</b>	
1.1	Pendahuluan	1
1.2	Latar Belakang Kajian	1
1.3	Penyataan Masalah	4
1.4	Matlamat dan Objektif Kajian	6
1.5	Skop Kajian	7
1.6	Kaedah Kajian	9
	1.6.1 Pembangunan Kerangka Kerja MalayAR	9
	1.6.2 Pembangunan Model Pelabelan Kelas Semantik	11
	1.6.3 Pembangunan Proses Capaian Maklumat Calon RP	11
1.7	Kepentingan Kajian	12
1.8	Organisasi Tesis	12
<b>BAB II</b>	<b>Sorotan Kesusasteraan</b>	
2.1	Pengenalan	14
2.2	Resolusi Anafora	15
	2.2.1 Jenis Anteseden dan Anafor	16
	2.2.2 Penentuan Pasangan yang Sesuai	18
	2.2.3 Pleonastik	23
	2.2.4 Rumusan	27

2.3	Kajian Terdahulu Resolusi Anafora	28
	2.3.1 Resolusi Anafora dalam Bahasa Inggeris	28
	2.3.2 Resolusi Anafora Selain Bahasa Inggeris	34
	2.3.3 Rumusan	43
2.4	Resolusi Anafora dalam Bahasa Malaysia	43
2.5	Pendekatan Resolusi Anafora	45
	2.5.1 Pendekatan Berasaskan Pengetahuan	48
	2.5.2 Pendekatan Pengetahuan Terhad	49
	2.5.3 Pendekatan Pembelajaran Mesin	55
	2.5.4 Rumusan	56
2.6	Pelabelan Kelas Semantik	56
	2.6.1 Gazetir	58
	2.6.2 Rumusan	59
2.7	Bahasa Melayu	59
	2.7.1 Kata Nama	59
	2.7.2 Kata Ganti Nama	63
	2.7.3 Enklitik <i>nya</i>	65
2.8	Perbincangan	67
	2.8.1 Penggunaan KGN	67
	2.8.2 Peralatan Sokongan dalam BM	69
	2.8.3 Pendekatan Penyelesaian	70
2.9	Kesimpulan	70
<b>BAB III</b>	<b>KAEDAH KAJIAN</b>	
3.1	Pengenalan	72
3.2	Fasa Pembangunan Kerangka Kerja	75
	3.2.1 Sorotan Kesusasteraan	75
	3.2.2 Penentuan Reka Bentuk Kerangka Kerja	76
3.3	Fasa Pembangunan Model Pelabelan Kelas Semantik	77
	3.3.1 Penentuan Kelas Semantik yang Sesuai	77
	3.3.2 Parameter untuk Setiap Kelas	80
	3.3.3 Pembangunan Gazetir	80
	3.3.4 Pembangunan Perkataan Saringan	80
	3.3.5 Pembangunan Petua	81
3.4	Fasa Pembangunan pengekrak Maklumat Calon RA	81
3.5	Fasa Pengujian dan Penilaian	82
	3.5.1 Data Pengujian dan Tanda Aras	83
	3.5.2 Komponen yang Diuji	85
	3.5.3 Peralatan Ujian	85
	3.5.4 Pengukur Penilaian	86

3.6	Kesimpulan	87
-----	------------	----

**BAB IV KERANGKA KERJA RESOLUSI ANAFOR UNTUK ARTIKEL BAHASA MELAYU (MALAYAR)**

4.1	Pengenalan	89
4.2	Isu Resolusi Anafora dalam BM	89
4.2.1	Jenis Kata Ganti Nama	90
4.2.2	Status Sosiologi	91
4.2.3	Penggunaan <i>-nya</i>	92
4.2.4	Jenis Rujukan	93
4.2.5	Kata Nama Khas	94
4.2.6	Pleonastik	95
4.2.7	Perkataan yang Mempunyai Sama Makna	97
4.2.8	Penghurai Separa SVO	97
4.2.9	Jarak Calon	98
4.3	Definisi Istilah	98
4.4	Kerangka Kerja MalayAR	100
4.4.1	Pembangunan Model Pelabelan Kelas Semantik	104
4.4.2	Ekstrak Maklumat Calon RA	110
4.4.3	Resolusi Anafora untuk Artikel Bahasa Melayu	123
4.5	Rumusan	134

**BAB V HASIL KAJIAN DAN PERBINCANGAN**

5.1	Pengenalan	136
5.2	Struktur Pengujian	136
5.2.1	Pembangunan Prototaip MalayAR	137
5.2.2	Penetapan Pengujian	139
5.3	Hasil Kajian Pemprosesan Calon Anafor Dan Anteseden	139
5.3.1	Hasil Kajian Pengecaman Entiti Nama	140
5.3.2	Hasil Kajian Pengecaman KGN <i>nya</i>	141
5.4	Hasil Kajian Kerangka Kerja Konsep MalayAR	142
5.4.1	Hasil Kajian Capaian Calon Anafor	143
5.4.2	Hasil Kajian MalayAR	145
5.4.3	Hasil Pendekatan Lain	147
5.4.4	Perbandingan Hasil Kajian MalayAR dan Pendekatan Lain	152
5.5	Kesimpulan	155

**BAB VI KESIMPULAN**

6.1	Pendahuluan	156
-----	-------------	-----



6.2	Sumbangan Kajian	156
6.3	Kelemahan dan Cadangan Kajian Masa Hadapan	159
6.4	Penutup	160

#### LAMPIRAN

A	Analisis Penentuan Status Sosiologi	171
B	Senarai Perkataan Kelas Semantik Untuk KNA, KK dan Ktgs	172
C	Perkataan Saringan	173
D	Gazetir	174
E	Artikel yang Digunakan dalam Kajian	177
F	Contoh Artikel yang Dianotasi	180
G	Analisis Penentuan Kelas Semantik	184
H	Analisis Penentuan Parameter Setiap Kelas Semantik	187
I	Senarai Set Kelas Semantik $AS_i$ Mengikut Kelas Kata	191
J	Senarai Gelaran Kehormat dan Kerabat	192
K	Sebahagian Hasil Pemprosesan Jenis Penggunaannya	194

## SENARAI JADUAL

No Jadual		Halaman
1.1	Kelas semantik	7
2.1	Ringkasan kajian terdahulu	22
2.2	Hasil ujian Boyd et al. (2005) berbanding Evans (2001a)	26
2.3	Ringkasan kajian dalam BI	31
2.4	Hasil perbandingan	38
2.5	Ringkasan kajian RA untuk bahasa selain BI	40
2.6	Perbandingan jenis pendekatan dalam RA	47
2.7	Perbandingan pendekatan pengetahuan terhad	49
2.8	Senarai imbuhan KNA dan contoh	62
2.9	Senarai KGN-O	64
2.10	Penggunaan KGN bagi setiap jenis sosiologi pengguna	65
2.11	Contoh penggunaan sebagai KGN dan pleonastik (penekan, pembenda)	66
2.12	Perbezaan antara pelbagai bahasa	68
4.1	Pecahan penggunaan KGN	91
4.2	Istilah yang digunakan dalam rumus dan algoritma	98
4.3	Faktor MalayAR	131
4.4	Penentuan skor ( $S_j$ )	133
5.1	Capaian dan pengelasan entiti nama	140
5.2	Hasil penentuan jenis rujukan KGN <i>nya</i>	142
5.3	Hasil capaian calon anafor	144
5.4	Hasil pengujian kadar kejayaan MalayAR	145
5.5	Hasil kadar kejayaan dan pengukuran perbandingan.	153

## SENARAI RAJAH

No. Rajah		Halaman
1.1	Ringkasan keseluruhan kaedah kajian	10
2.1	Contoh penghurai pohon	20
2.2	Penjelasan fitur [+3]	45
2.3	Ilustrasi pendekatan MOA	50
2.4	Gambaran prosedur MARS	52
2.5	Hubungan dan proses komponen RAP	52
2.6	Kerangka kerja ARN	54
2.7	Pecahan kata nama dalam BM	60
2.8	Pecahan kata ganti nama	63
3.1	Gambaran keseluruhan kerangka metodologi kajian.	74
3.2	Analisis penentuan kelas semantik	78
3.3	Faktor bagi menentukan kelas orang.	79
3.4	Contoh kandungan makna dalam kamus yang digunakan	81
3.5	Model pengujian dan penilaian	83
3.6	Contoh artikel berita	84
4.1	Graf peratusan penggunaan KGN dalam artikel	91
4.2	Pengecaman perkataan <i>nya</i>	96
4.3	Penentuan <i>nya</i> sebagai pleonastik	96
4.4	Kerangka kerja MalayAR	101
4.5	Senibina MalayAR	102
4.6	Hubungan faktor penentuan kelas semantik orang	106
4.7	Set petua penentuan kelas semantik bagi KNK	109
4.8	Penentuan kelas semantik bagi KNA.	110
4.9	Algoritma pemprosesan artikel MalayAR.	111

4.10	Antara muka pemrosesan artikel (hasil pemrosesan artikel)	112
4.11	Algoritma pemecahan ayat dalam artikel tidak berstruktur	113
4.12	Algoritma token untuk setiap ayat dalam setiap perenggan ( $a \subset p$ )	115
4.13	Paparan pemecahan perkataan	116
4.14	Algoritma pelabelan kelas kata	117
4.15	Algoritma penentuan kelas semantik untuk KNK	119
4.16	Pengasingan perkataan <i>nya</i> daripada kata hos <i>kh</i>	121
4.17	Petua penentuan jenis rujukan enklitik <i>nya</i>	122
4.18	Algoritma penentuan status sosiologi.	123
4.19	Algoritma capaian calon anafor KGN <i>nya</i>	125
4.20	Algoritma menentu KGN <i>mereka</i> sebagai calon anafor	126
4.21	Algoritma pemilihan calon anafora <i>beliau, baginda, dia</i> atau <i>ia</i>	127
4.22	Algoritma pemilihan calon anteseden berdasarkan calon anafor yang dipilih	129
5.1	Carta alir modul prototaip	138
5.2	Perbandingan hasil dapatan MalayAR dengan kepakaran manusia	146
5.3	Hasil RA di antara pendekatan Mitkov berbanding data penanda aras (pakar).	149
5.4	Perbandingan pendekatan Mitkov diaplikasi ke dalam BM dan pendekatan hasil kajian	150
5.5	Pendekatan Lappin dan Leass diaplikasikan ke atas bahasa selain Inggeris	151
5.6	Kadar kejayaan yang dihasilkan dengan menggunakan pendekatan Lappin dan Mitkov ke atas BM dan BI	152
5.7	Perbandingan kadar kejayaan MalayAR dengan pendekatan Mitkov dan Lappin	154

## SENARAI SINGKATAN

**Penggunaan dalam Bahasa Inggeris****Penggunaan dalam Bahasa Melayu**

AR	Anaphora Resolution	RA	Resolusi anafora
PN	Proper Nouns	KNK	Kata Nama Khas
NLP	Natural Language Processing	PBT	Pemprosesan Bahasa Tabii
Malay	Malay Language	BM	Bahasa Melayu
V	Verb	KK	Kata Kerja
N	Noun	KK	Kata Nama
F	Function word	Ktgs	Kata tugas
A	Adjective	KA	Kata Adjektif
Prn	Pronouns	KGN	Kata ganti nama
KP	Knowledge Poor	PT	Pengetahuan Terhad
WK	World-knowledge	PI	pengetahuan tabii

**SENARAI ISTILAH**

<b>Bahasa Melayu</b>	<b>Bahasa Inggeris</b>
Berasaskan Petua	Rule-based
Wacana	Discourse
Pengetahuan Terhad	Knowledge Poor
Sukatan Kedudukan Ketara	Salience Measures
Gazetir	Gazetteer
Kelas semantik	Semantic class
Ketepatan	Precision
Ukuran-f	F-measure
pemecah perkataan	Token
Penghurai	Parser
Pengukuran ciri khas	Salience measures
perintah-C	C-Command
keserasian semantik	Semantic agreement
Pengetahuan tabii	World knowledge

## RUJUKAN

- A.Akilandeswari & Devi, S. L. 2013. Conditional Random Fields based Pronominal Resolution in Tamil. . *International Journal on Computer Science and Engineering (IJCSE)* 5(6): 567-577.
- Agarwal, S., Srivastava, M., Agarwal, P. & Sanyal, R.: Anaphora Resolution in Hindi Documents. International Conference of Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. IEEE, Beijing, China (2007) 452 - 458
- Akilandeswari, A. & Sobha, L. D. 2012. Resolution for Pronouns in Tamil Using CRF. *The Workshop on Machine Translation and Parsing in Indian Languages*, hlm. 103-111.
- Alfred, R., Leong, L. C., On, C. K., Anthony, P., Fun, T. S., Razali, M. N. B. & Hijazi, M. H. A. 2013. A Rule-Based Named-Entity Recognition for Malay Articles. Dlm. Motoda, H., Wu, Z., Cao, L., Zaiane, O., Yao, M. & Wang, W. (pnyt.). *Advanced Data Mining and Applications*, hlm. 288-299. Springer Berlin Heidelberg.
- Alfred, R., Mujat, A. & Obid, J. H. 2013. *A Ruled-Based Part of Speech (RPOS) Tagger for Malay Text Articles*. Selamat, A., Nguyen, N. T. & Haron, H. Lecture Notes in Computer Science. Ed. Springer, Heidelberg: Springer Berlin Heidelberg.
- Andrei, P.-B. & Isabelle, R. 1997. Cooperation between pronoun and reference resolution for unrestricted texts. Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts. Madrid, Spain,
- Antonio, F., ndez, Jes & s, P. 2000. A computational approach to zero-pronouns in Spanish. Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. Hong Kong,
- Anuar, A. J. H. 1999. Ianya Benar. Persidangan Antarabangsa Pengajian Melayu/Indonesia : Ancangan Pembingkas Berdaya Cipta. Prince of Songkla University, Pattani, 21-23 Jun 1999.
- Barbu, C. 2003. Bilingual Pronoun Resolution: Experiments in English and French. University of Wolverhampton.
- Bayomi, M., Levacher, K., Ghorab, M. R., Lavin, P., O'Connor, A. & Lawless, S. 2016. Towards Evaluating the Impact of Anaphora Resolution on Text Summarisation from a Human Perspective. Dlm. Métais, E., Meziane, F., Saraee, M., Sugumaran, V. & Vadera, S. (pnyt.). *Natural Language Processing and Information Systems*, hlm. 187-199. Switzerland: Springer International Publishing.

- Bergsma, S. 2005. Automatic Acquisition of Gender Information for Anaphora Resolution Advances in Artificial Intelligence. Dlm. (pnyt.). hlm. 135-137. Springer Berlin / Heidelberg.
- Bergsma, S. & Yarowsky, D. 2011. NADA: A Robust System for Non-referential Pronoun Detection. Dlm. Hendrickx, I., Devi, S. L., Branco, A. & Mitkov, R. (pnyt.). *8th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2011*, hlm. 12-23. Faro, Portugal: Springer Berlin Heidelberg.
- Bethard, S. & Martin, J. H. 2006. Identification of event mentions and their semantic class. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, hlm. 146-154.
- Boyd, A., Gegg-Harrison, W. & Byron, D. 2005. Identifying Non-Referential it: A Machine Learning Approach Incorporating Linguistically Motivated Patterns. *ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, hlm. 40-47.
- Brown, P., Cocke, J., Pietra, S. D., Pietra, V. D., Jelinek, F., Mercer, R. & Roossin, P.: A Statistical Approach to Language Translation. In: Vargha, D. (ed.): *The 12th Conference on Computational Linguistics, Vol. 1*. Association for Computational Linguistics, Budapest, Hungary (1988) 71-76
- Budi, I., Bressan, S., Hasibuan, G. W. Z. A. & Nazief, B. A. A. 2005. Named Entity Recognition for the Indonesian Language: Combining Contextual, Morphological and Part-of-Speech Features into a Knowledge Engineering Approach. *Lecture Notes in Artificial Intelligence*, hlm.
- Carbonell, J. G. & Brown, R. D. 1988. Anaphora Resolution: A Multi-Strategy Approach. *Proceedings of the 12. International Conference on Computational Linguistics COLING'88*, hlm. 96-101.
- Chinatsu, A. & Douglas, M. 1993. A language-independent anaphora resolution system for understanding multilingual texts. *Proceedings of the 31st annual meeting on Association for Computational Linguistics*, hlm. 156-163.
- Chinatsu, A. & Scott, B. 1996. Applying machine learning to anaphora resolution. *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, hlm. 302-314.
- Converse, S. P. 2006. *Pronominal Anaphora Resolution in Chinese*. Computer and Information Science, University of Pennsylvania.
- Cuevas, R. R. M. & Paraboni, I. 2008. A Machine Learning Approach to Portuguese Pronoun Resolution. Dlm. Geffner, H., Prada, R., Alexandre, I. M. & David, N. (pnyt.). *Advances in Artificial Intelligence – IBERAMIA 2008*, hlm. 262-271. Lisbon, Portugal: Springer Berlin Heidelberg.
- Cui, Y., Hu, Q., Pan, H. & Hu, J. 2006. Zero Anaphora Resolution in Chinese Discourse. Dlm. (pnyt.). *Lecture Notes in Computer Science*, hlm. 245 – 248. Berlin Heidelberg: Springer-Verlag.



- Dagan, I. & Itai, A. 1990. Automatic Processing of Large Corpora for the Resolution of Anaphora References. *The 13th International Conference on Computational Linguistics*, hlm. 330-332.
- Dali, L., Rusu, D., Fortuna, B., Mladenici, D., Grob, M. & elnik. 2009. Question Answering Based on Semantic Graphs. Madrid, Spain, April 2009.
- Denber, M. 1998. Automatic Resolution of Anaphora in English. Eastman Kodak Co. Imaging Science Division.
- Denis, P. 2007. New Learning Models For Robust Reference Resolution. PhD, Faculty of the Graduate School, University of Texas at Austin.
- Derczynski, L., Maynard, D., Rizzo, G., Erp, M. v., Gorrell, G., Troncy, R. e., Petrak, J., Bontcheva, K. & A.Akilandeswari. 2014. Analysis of Named Entity Recognition and Linking for Tweets.
- Dewan Bahasa dan Pustaka: Tesaurus Bahasa Melayu Dewan Edisi Baharu. In: Mashitah binti Taharin, Rahmah binti Ja'afar & Shukor, N. b. A. (eds.): Tesaurus Bahasa Melayu Dewan Edisi Baharu. Dawama Sdn. Bhd, Kuala Lumpur (2010)
- Dewan Bahasa Pustaka. Pusat Rujukan Persuratan Malayu. *Khidmat Nasihat*. <http://prpm.dbp.gov.my/> [20 December 2010]
- Dimitrov, M., Bontcheva, K., Cunningham, H. & Maynard, D. 2002. A Light-weight Approach to Coreference Resolution for Named Entities in Text. *Fourth Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, hlm.
- Elghamry, K., Al-Sabbagh, R. & El-Zeiny, N. 2007. Arabic Anaphora Resolution Using the Web as Corpus. The Seventh Conference on Language Engineering. Cairo, Egypt.,
- Eugene, C. 2000. A Maximum-entropy-inspired Parser. *1st North American Chapter of the Association for Computational Linguistics Conference*, hlm. 132--139.
- Eugene, C. & Micha, E. 2009. EM works for pronoun anaphora resolution. Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Athens, Greece,
- Evans, R. 2001a. Applying Machine Learning Toward an Automatic Classification of It. *Literary and Linguistic Computing* 16(1): 45-57.
- Evans, R.: Mitkov's Anaphora Resolution System. Research Group in Computational Linguistics (2001b)
- Fazal Mohamed, M. S. 2006. Pengklitikan enklitik "-nya" pada kata kerja: aplikasi Teori Kuasaan dan Tambatan (Chomsky, 1986). *Jurnal Bahasa* 6(3):

- Fei, L. & Shuicai, S. 2008. Chinese Pronominal Anaphora Resolution Based on Conditional Random Fields. *2008 International Conference on Computer Science and Software Engineering*, hlm. 731-734.
- Ge, N. 2000. An Approach to Anaphoric Pronouns. Computer Science, Brown University.
- Green, R. & Dorr, B. J. 2005. Frame Semantic Enhancement of Lexical-Semantic Resources. *ACL-SIGLEX Workshop on Deep Lexical Acquisition*, hlm. 57-66.
- Hernández, I. Z. 2007. Demonstrative Pronouns In Spanish: A Discourse-Based Study. Department of Linguistics, Ohio State University.
- Hideki, I. & Tsutomu, H. 2003. Japanese zero pronoun resolution based on ranking rules and machine learning. Proceedings of the 2003 conference on Empirical methods in natural language processing - Volume 10.
- Hiroshi, N. 1997. Automatic extraction of rules for anaphora resolution of Japanese zero pronouns from aligned sentence pairs. Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts. Madrid, Spain,
- Hobbs, J. R. 1976. Pronoun Resolution. Research Report. City University of New York.
- Hobbs, J. R. 1986. Resolving pronoun references. Dlm. (pnyt.). *Readings in natural language processing*, hlm. 339-352. Morgan Kaufmann Publishers Inc.
- Holen, G. I. 2006. Automatic Anaphora Resolution for Norwegian (ARN). Department of Linguistics and Scandinavian Studies, UNIVERSITY OF OSLO.
- Holen, G. I. 2007. Automatic Anaphora Resolution for Norwegian (ARN). Dlm. Branco, A. (pnyt.). *Anaphora: Analysis, Algorithms and Applications: 6th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2007, Lagos, Portugal, March 29-30, 2007. Selected Papers*, hlm. 151-166. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hwee Tou, N., Yu, Z., Robert, D. & Mary, G. 2005. A machine learning approach to identification and resolution of one-anaphora. Proceedings of the 19th international joint conference on Artificial intelligence. Edinburgh, Scotland,
- Iida, R., Inui, K. & Matsumoto, Y. 2005. Anaphora resolution by antecedent identification followed by anaphoricity determination. *ACM Transactions on Asian Language Information Processing (TALIP)* 4(4): 417-434.
- Ingria, R. J. P. & Stallard, D. 1989. A Computational Mechanism for Pronominal Reference. *The 27th Annual Meeting of the ACL*, hlm. 262-271.

- James, F., Ken, B. & Bruce, P. 2005. Indirect anaphora resolution as semantic path search. Proceedings of the 3rd international conference on Knowledge capture. Banff, Alberta, Canada,
- Johnson, B. 2010. A Rule-Based System for Identifying Pleonastic 'It'. Master, Department of Computer Science and Electrical Engineering, University of Maryland.
- Joseph McKendrick. 2012. The Post-Relational Reality Sets In: 2011 Survey on Unstructured Data. Analyst Report. Unisphere Research, a Division of Information Today, Inc.
- Kamune, K. P. & Agrawal, A. 2015. Hybrid Approach to Pronominal Anaphora Resolution in English Newspaper Text. *International Journal Intelligent Systems and Applications* 2015(02): 56-64.
- Kawahara, D. & Kurohashi, S. 2004. Zero Pronoun Resolution Based on Automatically Constructed Case Frames and Structural Preference of Antecedents. Dlm. Su, K.-Y., Tsujii, J. i., Lee, J.-H. & Kwong, O. Y. (pnyt.). *Natural Language Processing – IJCNLP 2004*, hlm. 2-21. Springer-Verlag Berlin Heidelberg: Springer Berlin Heidelberg.
- Kennedy, C. & Boguraev, B. 1996. Anaphora for everyone: Pronominal anaphora resolution without a parser. *16th International Conference on Computational Linguistics*, hlm.
- Koga, T., Wu, H. & Furugori, T. 1999. Determining the Antecedent of Noun Phrase Containing the Determiner KONO or SONO in Japanese. *The 13th Pacific Asia Conference on Language, Information and Computation*, hlm. 251-259.
- Kong, F. & Zhou, G. 2010. A Tree Kernel-Based Unified Framework for Chinese Zero Anaphora Resolution. *Conference on Empirical Methods in Natural Language Processing*, hlm. 882-891.
- Kong, F. & Zhou, G. 2011. Improve Tree Kernel-Based Event Pronoun Resolution with Competitive Information. Twenty-Second International Joint Conference on Artificial Intelligence. Centre Convencions Internacional Barcelona, 16–22 July 2011.
- Lakhmani, P., Singh, S. & Mathur, P. 2014. Gazetteer Method for Resolving Pronominal Anaphora in Hindi Language. *International Journal of Advances in Computer Science and Technology* 3(3): 51-57.
- Lappin, S. & Leass, H. J. 1994. An algorithm for pronominal anaphora resolution. *Comput. Linguist.* 20(4): 535-561.
- Lejtovicz, K. E. & Kardkovács, Z. T. 2007. Anaphora Resolution. 8th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics. Hungary, 15-17 Nov 2007.

- Li, D., Miller, T. & Schuler, W.: A pronoun anaphora resolution system based on factorial hidden Markov models. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Vol. 1. Association for Computational Linguistics, Portland, Oregon (2011) 1169-1178
- Liang, T. & Wu, D.-S. 2003. Automatic Pronominal Anaphora Resolution in English Texts. *ROCLING.*, hlm. 21-40.
- Liddy, E. D. R. 1990. Anaphora in natural language processing and information retrieval. *Inf. Process. Manage.* 26(1): 39-52.
- Liu, D., McVeety, S., Prasad, R. & Natarajan, P. 2008. Semi-Supervised Topic Classification For Low Resource Languages. International Conference of Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. Las Vegas, NV, 31 March 2008 - 4 April 2008.
- Liu, Y. & Ren, F.-j. 2011. Japanese named entity recognition for question answering system International Conference on Cloud Computing and Intelligence Systems (CCIS). Beijing, China,
- McCord, M. C. 1990. Slot Grammar: A System for Simpler Construction of Practical Natural Language Grammars. International Symposium on Natural Language and Logic.
- Mitkov, R. 1994. An integrated model for anaphora resolution. Proceedings of the 15th conference on Computational linguistics - Volume 2. Kyoto, Japan,
- Mitkov, R. 1996. Pronoun Resolution: The Practical Alternative. Discourse Anaphora and Anaphor Resolution (DAARC). Lancaster, UK,
- Mitkov, R. 1997. Factors in anaphora resolution: they are not the only things that matter: a case study based on two different approaches. Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts. Madrid, Spain,
- Mitkov, R. 1999. Anaphora Resolution: The State of The Art. University of Wolverhampton.
- Mitkov, R. 2002a. An approach in focus: Mitkov's robust, knowledge-poor algorithm. Dlm. (pnyt.). *Anaphora Resolution*, hlm. 220. Great Britain: Pearson Education.
- Mitkov, R. 2002b. Linguistic Fundamentals. Dlm. (pnyt.). *Anaphora Resolution*, hlm. 220. Great Britain: Pearson Education.
- Mitkov, R. 2002c. Outstanding Issues. Dlm. (pnyt.). Mitkov, R. *Anaphora Resolution*, hlm. 192-197. Great Britain: Pearson Education limited.

- Mitkov, R. 2002d. The Past: Work in the 1960s, 1970s and 1980s. Dlm. Mitkov, R. (pnyt.). Mitkov, R. *Anaphora Resolution*, hlm. 68-92. Great Britain: Pearson Publication.
- Mitkov, R. 2002e. The Present: Knowledge-poor and corpus approaches in the 1990s and beyond. Dlm. (pnyt.). *Anaphora Resolution*, hlm. 95-129. Great Britain: Pearson Education.
- Mitkov, R. 2002f. The Process of Automatic Anaphora Resolution. Dlm. (pnyt.). *Anaphora Resolution*, hlm. 28-52. Great Britain: Pearson education.
- Mitkov, R., Belguith, L. & Stys, M. 1998. Multilingual robust anaphora resolution. *In Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*, hlm. 7-16.
- Mitkov, R., Evans, R., Orasan, C., Ha, L. A. & Pekar, V. 2007. Anaphora Resolution: To What Extent Does It Help NLP Applications? Dlm. (pnyt.). *Anaphora: Analysis, Algorithms and Applications*, hlm. 179-190. Springer Berlin / Heidelberg.
- Mitkov, R., Richard, E. & Constantin, O. 2002. A New, Fully Automatic Version of Mitkov's Knowledge-Poor Pronoun Resolution Method. *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing*.
- Mohd Juzaidin Ab Aziz. 2008. Pola Grammar for Automated Marking of Malay Short Answer Essay-Type Examination. *Computer Science and Information Technology*, Universiti Putra Malaysia.
- Mohd Juzaidin Ab Aziz, Fatimah Dato' Ahmad, Abdul Azim Abdul Ghani & Ramlan Mahmud. 2009. Automated Marking System for Short Answer examination (AMS-SAE). *Industrial Electronics & Applications*. ISIEA 2009.
- Mohd Pouzi Hamzah. 2006. Frasa Dan Hubungan Semantik Dalam Perwakilan Pengetahuan: Kesan Terhadap Keberkesanan Capaian Dokumen Melayu. *Fakulti Teknologi dan Sains Maklumat*, Universiti Kebangsaan Malaysia.
- Mohd Zakree Nazri, Siti Mariyam Shamsuddin, Azuraliza Abu Bakar & Salwani Abdullah. 2011. A hybrid approach for learning concept hierarchy from Malay text using artificial immune network. *Natural Computing: an international journal* 10(1): 275-304.
- Mohamad, R. 2013. Kerangka Kerja Sistem Sokongan Keputusan Pengurusan Tender Berasaskan ontologi. *Fakulti Teknologi dan Sains Maklumat*, Universiti Kebangsaan Malaysia.
- Muñoz, R., Saiz-Noeda, M. & Montoyo, A. 2002. *Semantic Information in Anaphora Resolution*. *Lecture Notes in Computer Science*. Ed. Heidelberg: Springer Berlin / Heidelberg.

- Muhamad Taufik Abdullah, Fatimah Ahmad, Ramlan Mahmood and Tengku Mohd Tengku Sembok. 2009. Rules Frequency Order Stemmer for Malay Language. *International Journal of Computer Science and Network Security* 9(2): 465.
- Muhammad, A.-M. 2011. Automatic Detection of Arabic Non-Anaphoric Pronouns for Improving Anaphora Resolution. *Journal of ACM Transactions on Asian Language Information Processing (TALIP)* 10(1): 1-11.
- Nik Safiah, K. 1995. *Malay Grammar for Academics and Professionals*. Ed. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Nik Safiah, K., Farid, M. O., Hashim, H. M. & Abdul Hamid, M. 2008. *Tatabahasa Dewan*. Ed. 3. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Niraula, N. & Rus, V. 2014. A Machine Learning Approach to Pronominal Anaphora Resolution in Dialogue Based Intelligent Tutoring Systems. Dlm. Gelbukh, A. (pnyt.). *Computational Linguistics and Intelligent Text Processing*, hlm. 307-318. Springer Berlin Heidelberg.
- Nøklestad, A. 2009. A Machine Learning Approach to Anaphora Resolution Including Named Entity Recognition, PP Attachment Disambiguation, and Animacy Detection. Faculty of Humanities, The University Of Oslo.
- Nomoto, H. 2011. Teori tambatan dalam bahasa Melayu. *Southeast Asian studies, Tokyo University of Foreign Studies* 16(1-16).
- Nor Hashimah, J. 2006a. Kata Ganti Nama 'IA' dan 'Ianya' dalam Korpus Melayu: Kajian rangka Rujuk Silang. Dlm. (pnyt.). *Nahu Praktis Bahasa Melayu*, hlm. Bangi, Selangor: UKM.
- Nor Hashimah, J. 2006b. Kata ganti nama ia dan ianya. Dlm. Zaharani, A. (pnyt.). *Aspek Nahu Praktis Bahasa Melayu*, hlm. 232-248. Bandar Baru Bangi: Penerbit Universiti Kebangsaan Malaysia.
- Omar, A. 2009. *Nahu Melayu Mutakhir*. Ed. 5. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Omar, N., Hanna, P. & McKeivitt, P. 2006. Semantic analysis in the automation of ER modelling through natural language processing. *Computing & Informatics, 2006. ICOCI '06. International Conference on*, hlm. 1-5.
- Paice, C. D. & Husk, G. D. 1987. Towards the automatic recognition of anaphoric features in English text: The impersonal pronoun "it". *Computer Speech and Language* 2(
- Pantel, P. & Ravichandran, D. 2004. Automatically Labeling Semantic Classes. *Proceedings of Human Language Technology / North American Association for Computational Linguistics (HLT/NAACL-04)*, hlm. 321-328.

- Qinan, H. 2008. A Corpus-Based Study on Zero Anaphora Resolution In Chinese Discourse. PhD Thesis, Department of Chinese, Translation and Linguistics, City University of Hong Kong.
- Qiu, L., Kan, M.-Y. & Chua, T.-S. 2004. A Public Reference Implementation of the RAP Anaphora Resolution Algorithm. *Proceedings of the Language Resources and Evaluation Conference 2004 (LREC 04)*, hlm. 291-294.
- Rabiah, A. K. 2008. Question Answering for Reading Comprehension Using Logical Inference Model. Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia.
- Rich, E. & LuperFoy, S. 1988. An Architecture for Anaphora Resolution. *Second Conference on Applied Natural Language Processing (ANLP-2)*, hlm.
- Ryohei, S., Daisuke, K. & Sadao, K. 2008. A fully-lexicalized probabilistic model for Japanese zero anaphora resolution. *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*. Manchester, United Kingdom,
- Sam, C.-S. 1991. Automatic Acquisition of Proper Noun Meanings. *Proceedings of the 6th International Symposium on Methodologies for Intelligent Systems*.
- Samuel, W. K. C. & Benjamin, K. T. S. 1999. Semantic Inference for Anaphora Resolution: Toward a Framework in Machine Translation. *Machine Translation* 14(3/4): 163-190.
- Sarkar, S., Madasu, V. R., SM, B. M. & Rao, S. V. 2015. NLP Algorithm Based Question and Answering System. *Proceedings of the 2015 Seventh International Conference on Computational Intelligence, Modelling and Simulation.*, hlm. 97--101.
- Seki, K., Fujii, A. & Ishikawa, T. 2002. A probabilistic method for analyzing Japanese anaphora integrating zero pronoun detection and resolution. *Proceedings of the 19th international conference on Computational linguistics*. Taipei, Taiwan,
- Shalom, L. & Michael, M. 1990. A syntactic filter on pronominal anaphora for Slot Grammar. *Proceedings of the 28th annual meeting on Association for Computational Linguistics*. Pittsburgh, Pennsylvania,
- Sikdar, U. K., Ekbal, A. & Saha, S. 2016. A generalized framework for anaphora resolution in Indian languages. *Knowledge-Based Systems* Articles in Press.
- Singh, S., Lakhmani, P., Mathur, P. & Morwa, S. 2014. Analysis of Anaphora Resolution System for English Language. *International Journal on Information Theory (IJIT)* 3(2):
- Sobha, L. 2007. Resolution of Pronominals in Tamil. *Computing: Theory and Applications, 2007. ICCTA '07. International Conference on*, hlm. 475-479.

- Stuckardt, R. 2007. Applying Backpropagation Networks to Anaphor Resolution. Dlm. (pnyt.). *Anaphora: Analysis, Algorithms and Applications*, hlm. 107-124. Springer Berlin / Heidelberg.
- Tapanainen, P. & Jarvinen, T. 1997. A non-projective dependency parser. *the 5th Annual Conference of Applied Natural Language Processing*, hlm. 64-71.
- Vicedo, J. L. & Ferrández, A. 2000. Applying Anaphora Resolution to Question Answering and Information Retrieval Systems. Dlm. (pnyt.). *Lecture Notes in Computer Science*, hlm. Heidelberg: Springer Berlin / Heidelberg.
- Wei, X., Zang, H. & Zhang, Q. 2008. Disambiguate Chinese Personal Pronoun Based on Semantic Structure. *IEEE International Conference in Granular Computing*, hlm. 644-648.
- Wu, D.-S. & Liang, T. 2009. Zero anaphora resolution by case-based reasoning and pattern conceptualization. *Expert Systems with Applications* 36(4): 7544-7551.
- Xiaofeng, Y., Jian, S. & Chew Lim, T. 2005. Improving pronoun resolution using statistics-based semantic compatibility information. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, hlm. 165-172.
- XUE, N., XIA, F., CHIOU, F.-D. & PALMER, M. 2005. The Penn Chinese TreeBank: Phrase structure annotation of a large corpus. *Natural Language Engineering* 11(02): 207--238.
- Yöndem, D. K. a. M. T. 2007. A Knowledge-poor Pronoun Resolution System for Turkish. *6th Discourse Anaphora and Anaphora Resolution Colloquium (DAARC)*, hlm.
- Zhang, Z. & Iria, J. 2009. A novel approach to automatic gazetteer generation using Wikipedia. Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources. Suntec, Singapore,